DEFINING HOSPITAL PEER GROUPS: WHY IS IT SO DIFFICULT?

Elisa Honório, Ana Lírio, Céu Mateus, Pedro Pita Barros, Maria do Céu Valente, Margarida Bentes
Ministry of Health, Portugal

Communications to: Maria do Céu Valente, Ministry of Health, Av. da República n. 61, 2nd floor, 1050 Lisbon, Portugal, Tel. +35.1.7925520, Fax +351.1.7925523, E-mail cvalente@igif.min-saude.pt

ABSTRACT
The objective of this paper is to present the background work underlying the definition of hospital peer groups. The groups are defined for financing purposes. The grouping procedure takes into account the fixed component of hospital expenditures, independently of production, quality of provision or efficiency. Classification of hospitals into groups means that we want to group together "similar" hospitals. The set of hospital characteristics was selected on the basis of previous studies performed on hospital grouping, and considering the available information as well as its quality. The hierachical cluster analysis chosen creates partitions which allow an easy merging of the groups created to the desired number of groups. Several agglomerative algorithms were tested, namely: the average distance between groups, the average distance intra groups, single linkage, complete linkage, centroid, median and ward method. Initially, there are n single groups (each observation in a group of one element only) and, for one, its most similar pair is picked to form a new group. The resulting groups were submitted to the panel of experts, which were asked to validate the resulting groups, bearing in mind the methodology used, the number of groups and their constitution. In order to compare the different methodologies used, and to measure the impact of changes suggested as well as of future political decisions, two measures were implemented: the minimum-variance type of measure (*vm*) and disjunction measures. Based on all the previous results, the choice of the panel of experts fell on coding using the same amplitude intervals and the creation of two groups for a set of outlier hospitals and three groups for the remaining population. A final, residual, group of hospitals includes hospitals presenting very specific characteristics that interfere in the homogeneity of the groups formed.
KEYWORDS: Portugal; Hospital Peer Groups; Hierachical Cluster Analysis; Minimum Variance Measure; Disjunction Measures.

NB: Tables indicated with roman numbers are not essential to the understanding of the text and can be found in the web page of the Journal.

INTRODUCTION

Since the implementation of DRGs in Portugal the need to group hospitals arose from the recognition that variations in production costs associated with teaching/research status and/or production structure (number of beds, human resources, specialities, etc.) were large and not related with efficiency or quality of care. This fact determined that the payment of DRG rates should be based on average production costs of hospital peer groups.

So far, the grouping methodology has been based on an existing administrative classification of hospitals (central and district). This classification was further readjusted by observation of outlier values of casemix index and cost per patient of individual hospitals. Five groups of hospitals were determined, based on these considerations. To each group was given a different financial treatment. The definition of these groups has been criticised by many hospitals, which claimed to have high cost structures, related to the role that they fulfil in the health system, that were not adequately accounted for by the variables used, namely, casemix index and cost per case.

The objective of this paper is to present the background work underlying the redefinition of hospital peer groups mainly for financing purposes. The goal was to develop a model that described, in a global and comprehensive way, the fixed component of the hospitals expenditures independently of production, quality of provision or efficiency. In our point of view, the variable component, associated with the production, is already reflected in the current funding model through DRG payments. Therefore, beyond the financial purposes, this new grouping will allow more accurate comparisons of institutions and the set up of reference values for performance, efficiency and quality benchmarking.

Statistical tools used to build the groups were based on the multivariate analysis – clusters analysis - and have the ability to distinguish information and highlight the underlying structure of the analysed objects, i.e., evidence similar subjects. The development of this work was composed by several stages, described below. Different sets of assumptions were considered, to evaluate the reliability and stability of the results.

Considering the limitations of statistical information as an adequate way of getting a good snapshot for every single hospital, moreover through such a few variables, the need of another kind of information was felt. Thus, a panel of experts in hospital management was brought in to evaluate results obtained from different approaches, at distinct stages, in a very interactive process, and adding what we called exogenous information.

The final groups obtained are substantially different from the ones based upon the administrative classification, what justifies the existing doubts concerning their validity for financing purposes.

DATA AND METHODOLOGY

Classification of hospitals into groups means that we want to group together "similar" hospitals. Thus implies the definition of a set of characteristics, over which we will build the grouping. The set of hospital characteristics was defined taking into account studies performed on hospital grouping, and considering the available information as well as its quality. The information constraints always play an important role in applications. Also here, the availability of information will dictate some options. The characteristics chosen to reflect the structure/complexity of institutions were dimension, diversity and casemix of care provided and available resources. Structure and complexity are not straightforward measurable therefore they had to be ascertained through a set of what we believe to be relevant characteristics.

The process of selecting variables was done having in mind the desire to minimise the influence of factors such as productivity, management of hospital resources, quality of care provided and efficiency, considering that these may vary widely in similar hospital structures.

Considering the universe of Portuguese hospitals (90 NHS institutions) it was our objective to investigate the underlying structure within the set of objects, in order to create a limited number of hospital groups (between five and eight). To achieve this, a multivariate analysis approach was performed, with emphasis on classification methods, allowing a more efficient organisation of the information and an easier assignment of new objects. In this kind of analysis, one of the major problems is usually the mixed nature of the data used (different units). To overcome that, all variables have to be converted into comparable units.

Previous to the present work there were two initial studies. In one of them, the numeric variables were categorised into intervals of equal length, respecting Sturges' rule which enables the relationship between the number of objects and the number of intervals that will be determined. In another one, scores (ranks) were assigned to objects' attributes, i.e., each variable is sorted by the observed values, starting from the lowest, which is given position number 1, to the highest, *n*, where *n* is the number of observations.

In spite of some agreement between both, the existing divergences were relevant enough to deserve further discussion. In order to understand what was causing such different results – differences of statistical methodology or differences in raw data – the first decision was to use the same matrix of raw data, and to keep both coding methodologies (same amplitude intervals and ranks) to control for potential distortion introduced by the conversion of data.

Considering the purpose, the available information, the variables included in previous studies and the panel of experts' opinion, the selected variables to group hospitals are shown in table 1.

Table 1. Selected variables

1. No. of beds	BED: translates a hospital's size, excluding the emergency service beds, new-borns cradles, recovery beds and intensive care and burning care beds.
2. No. of beds in special units	SPE_BEDS: in a way is a first approach to the casemix of the care provided, corresponds to the number of beds in the intensive unit and burning unit.
3. Casemix index	CMI: this index, which is fundamental for the determination of a hospital's budget, is based in the inpatient production measured through DRGs and it allow us to compare a hospital with all the others in terms of the patients' complexity and, consequently, in terms of resource consumption.
4. No. of physicians and nurses	PHYNUR: translates the personnel allocated to the hospital, it is a simple counting of the number of physicians and nurses working there.
5. No. of DRGs	FAN: counts the number of distinct DRG that the hospital presents and enables us to evaluate the diversity of the provided care.
6. Percentage of episodes of care grouped in "differentiated" surgical DRGs	DSDRGS: The existence of a surgical procedure is a variable, which influences the grouping of an episode of care in a DRG. The existence of a medical act of this nature is significant for the implications in costs, considering that it is connected with the use of the surgical theatre, recovery room and, eventually, after surgery intensive care. Therefore, it was selected a set of surgical DRGs (Annex 2) which are supposed to translate a larger consumption of resources in an inst itution, considering their proportion in the total inpatient production.
7. Type of diagnoses and therapeutic ancillaries	DTA: counts the types of diagnoses and therapeutic ancillaries existing in each institution and indicates the available resources and their diversity. For this study ten areas/types of ancillaries were considered: 1) pathologic anatomy; 2) angiography; 3) endoscopy; 4) immuno-hemotherapy; 5) lithotripsy; 6) nuclear medicine; 7) radiotherapy; 8) X-rays; 9) magnetic resonance imaging; 10) computerised axial tomography. The diagnoses and therapeutic ancillaries selected were, in their majority, considered as the more differentiated, when evaluating the existing resources and the care provided in the institution.

The source used to collect the data were the Statistics of the General Directorate of Health (1997), except for the oncology institutes, where the source was the National Institute of Statistics. The variables casemix index, percentage of episodes of care grouped in "differentiated" surgical DRGs and number of distinct DRGs were collected from DRGs national database, for the same year.

Finding outliers

Detection of outlier cases was performed by geometrical techniques, which allowed the representation of the subjects as points in a low-dimensional euclidian space, having the property that similars are presented close together. To the observer, the outlier data assessment becomes possible although there is the risk of discarding potential information.

Clustering of hospitals

There are several techniques to cluster and the option fell on the hierarchical one. This creates partitions which allow an easy merging of the groups created to the desired number of groups. Nevertheless, it should be noted that they are blind for the reallocation of objects that may have been "incorrectly" grouped at an earlier stage.

Classification methods require the replacement of the raw matrix data, where every individual is described by several attributes, for a matrix of dissimilarity values. In this preliminary step the accuracy of the chosen measure plays an important role.

The squared euclidean distance was applied to the data matrixes obtained from coding, to build the dissimilarity matrixes. Several agglomerative algorithms were tested, namely: the average distance between groups, the average distance intra groups, single linkage, complete linkage, centroid, median and ward method. These methods are iterative processes. Initially, there are n single groups (each observation in a group of one element only) and, for one, its most similar pair is picked to form a new group. Subsequently, the dissimilarities between the new group and the other groups are again recalculated (ex.: the minimum, the maximum, the mean,...).

To guarantee the stability of the groups it was decided that the final results would be achieved by the intersection of different agglomerative methods.

GROUPING HOSPITALS

In the first stage of the study the two maternities, the two orthopedic hospitals, the pneumology and infectious diseases and the ophthalmology hospital were not analysed because the panel of experts considered them to be very different from the rest of the population. Having as a basis a data matrix with a dimension (84 x 7) and after a simple descriptive analysis it was found that the population under analysis was, not surprisingly, heterogeneous.

Fifteen outlier hospitals were found using multidimensional scaling and principal co-ordinates analysis and were coincident in both techniques. This set of hospitals is, in a coherent way, distinct from the rest of the population; however, it is not homogeneous in its elements. Because of this, the group was studied apart from the rest although with the same statistical tools.

The comparison of coding methods results enables the evaluation of the distortion impact introduced into the original information. Using equal length intervals transformation, clusters are sensitive to the presence of subjects presenting behaviours distinct from the rest of the population. On the contrary, ranks turn the distances between subjects uniform, attenuating the real differences of raw data.

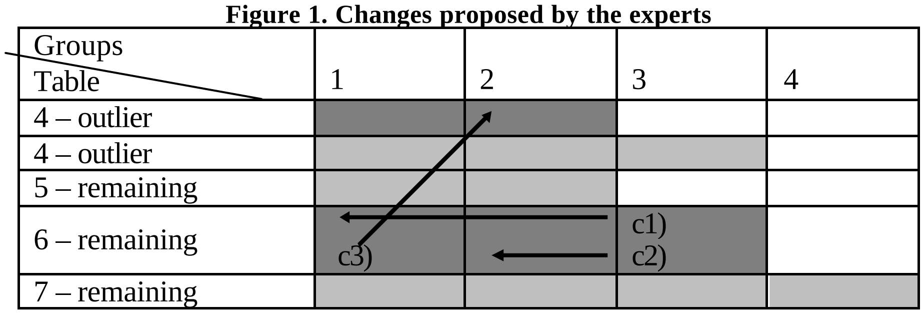
Therefore, a first proposal for groups was created and it was found that there was a high agreement between the different methodologies used to group hospitals. The resulting groups were submitted to the panel of experts, which were asked to validate the resulting groups, bearing in mind the methodology used, the number of groups and their constitution.

From tables V to VIII we can see that the experts suggestions originated the changes shown in figure 1.

In order to compare the different methodologies used, and to measure the impact of the changes suggested as well as of future political decisions, two measures were implemented: the minimum variance measure (*vm*) and disjunction measures. These measures intend to provide a feeling to how much statistical performance is gained/lost by ad-hoc adjustments. In our setting, this is a quite important issue. It is virtually impossible to reflect all heterogeneity of hospitals in a small number of groups, and the small set of characteristics does not fully reflect all differences across hospitals. Thus, the use of exogenous information should not be discarded. These measures give a way to evaluate the value of such

exogenous information, described as suggested changes in hospital groups. Note that these measures do not coincide with the objective functions associated with the methods used for grouping.

Figure 1. Changes proposed by the experts



Legend: The shadow means the number of groups present in each table.
– unselected table – selected tables

Minimum variance measure, *vm*, intends to ascertain the degree of variance of each grouping methodology. It should be noted that it is not surprisingly to obtain values of *vm* smaller than those in the original groups do for each of the methodologies. This occurs because cluster methods do not minimise the variance inside each group but rather attempt to reduce the dissimilarities between the elements of each group.

From the analysis of Table IX it is possible to observe a better performance of coding A in relation to coding B. In both methodologies, the better value of minimum variance measure occurs when changes respecting suggestions c1) and c2) are done.

Taking into account the suggestions of the experts, *vm* was calculated for methodology B with 6 groups (Table X).

Values from Tables IX and X are not comparable because the number of groups is different. However, it is possible to observe that the lowest value of *vm* is achieved, once again, when we do all the changes suggested but the one concerning H 89.

The second measure implemented, a disjunction measure, makes the split between two groups measurable and is called disjunction measure between group i and group j - D_{ij}.

When evaluating the results of this measure, in the different hypothesis, it was found that the change of group of H 63 had positive effects in the disjunction of the different groups. For the remaining changes, that was not verified.

Statistical evaluation of the ungrouped subjects

For the hospitals withdrawn from the population under analysis, several insertions in the groups obtained were simulated through discriminant analysis. This type of analysis assign new cases to one of the mutually exclusive groups previously defined, minimising the probability of a misclassification from a normal multivariate probabilistic model. Through the adjustment test of Kolmogorov-Smirnov, the Skewness and Kurtosis coefficients and graphical analysis (Q – Q plot and histograms), it was concluded that the variables did not come from a normal multivariate distribution. Thus, the results from the discriminant analysis were not reliable and moreover inconsistent.

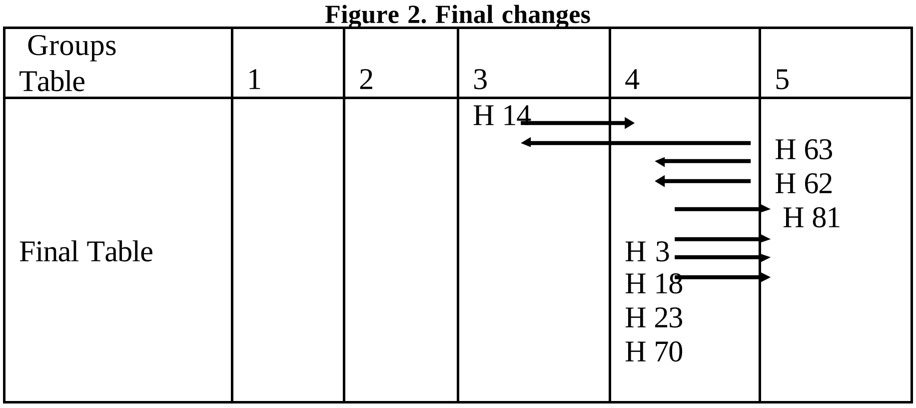
These partitions of the hospitals into a fixed number of disjoint groups, might not be the most accurate adjustment to the underlying data structure, it can even be inappropriate to oblige an object to belong to

only one group. That is what happens in the presence of objects, like the ungroupable hospitals, which probably have a degree of membership in several categories.

Final groups evaluation
Based on all the previous results, the choice of the panel of experts fell on coding using the same amplitude intervals and the creation of two groups for the outlier hospitals and three groups for the remaining population.

Nevertheless, after a final evaluation of the results obtained the final changes shown in figure 2 were done.

Figure 2. Final changes



Again, these groups were evaluated by the statistical measures described above and the results shown in table 2 were found:

Table 2. Minimum Variance Measure (*vm*)

Group	*Vm*	Contribution (%)	Original Vm	Change (%)
1	461,6	11,8%	461,6	0%
2	1512,6	38,8%	1512,6	0%
3	1089,2	27,9%	1060,7	2,6%
4	568,8	14,6%	659,2	-15,9%
5	265,6	6,8%	384,0	-44,6%
Grouping	3897,8	100,00%	4078,1	-4,6%

Thus, it is concluded that groups 2 and 3 are the most heterogeneous and that the apparent homogeneity of group 1 is due, mainly, to the fact that it includes only 3 hospitals. Once groups are ordered by the hospitals' dimension - except for groups 2 and 3 which were also ordered by casemix - it is possible to infer that a large dimension implies a higher heterogeneity within groups.

Analysing this table and considering a certain order among groups, it could be expected that, across lines, the value of the measures would rise and, across columns, a symmetric behaviour would occur. However, the disjunction between group 1 and 3 and between group 5 and 2 does not fulfil this expectation. This can be explained by the group's composition, for instance, hospitals belonging to groups 1 and 3 are all general hospitals and in group 2 we have some specialised institutions (table 3).

Table 3. Measure of disjunction (D_{ij})

D_{ij}	2	3	4	5
1	3,758	2,669	5,739	7,337
2		0,772	1,100	1,143
3			1,178	1,425
4				0,741

The differences between the actual groups and the administrative ones can be ascertained through the comparison of tables XII and XIII.

The statistical evaluation of the administrative groups is shown in table 4 and 5.

Table 4. Minimum Variance Measure (*vm*) for the administrative groups

Group	*Vm*	Contribution (%)
1	110,23	1,87%
2	3337,82	56,62%
3	425,48	7,22%
4	1615,37	27,40%
5	406,52	6,90%
Grouping	5895,42	100,00%

Table 5. Measure of disjunction (D_{ij}) for the administrative groups

D_{ij}	2	3	4	5
1	0,660	0,389	1,345	1,334
2		0,418	0,658	0,833
3			0,382	0,477
4				0,732

In spite of the fact that we have the same number of groups, values are not straightly comparable because subjects included in each group are not the same for the administrative partition and the final one. However, it is possible to notice that the variance of the administrative classification is greater than the variance of the final classification achieved in 51%. Moreover, it is also evident a larger disjunction between the final groups created than in the administrative ones.

CONCLUSION

The development of the present study respected the different stages required to achieve consistent and homogeneous hospital peer groups, namely selecting the variables which will be used in the model, as well as the discussion and validation of the results. The multivariate analysis performed clearly showed that the set of NHS hospitals includes very dissimilar realities not always comparable. According to the methods used and the chosen coding (same amplitude intervals) the existence of standardised groups was observed, as well as the existence of some unstable hospitals in the groups. There is still a residual group of hospitals presenting characteristics that interfere in the homogeneity of the groups formed.

The resulting hospital groups are expected to explain, to a large extent, variations between hospitals associated with their structure costs and their role in the health sector. In this respect, the new peer grouping will contribute to improve the performance of the current NHS DRG funding system

Nonetheless, we should not forget that hospitals are organisations in mutation and permanent adaptation. A hospital is by nature an open system that is influenced by and influences the environment in which it operates. Therefore, hospital peer grouping should be a dynamic work and the present results are fundamental for future enhancements of grouping methodologies, namely fuzzy sets. This technique is more adequate to classify objects that lay continuously in the multidimensional space, avoiding the fixed-boundary problem typical of conventional clustering approach. This is left for future work.

BIBLIOGRAPHY

1. Bentes, M.; J. Urbano and D. Hindle (1989) – Output-based funding in Portugal. Lisbon.
2. Gordon, A.D. (1981) – Classification. London, Chapman and Hall.
3. Johnson, R. and D. Wichern (1982) – Applied Multivariate Statistical Analysis. New Jersey, Prentice-Hall
4. Klecka, W. R. (1980) – Discriminant analysis. SAGE University Paper 19, USA, SAGE publications.
5. Lebert, L.; A. Morineau & K. M. Warwick (1989) – Multivariate descriptive statistical analysis, correspondence analysis and related techniques for large matrices. Canada, John Wiley & Sons.
6. Urbano, J.; M. Bentes and J. Vertrees (1991) – DRGs in Portugal. Lisbon.
7. Vertrees, J. et al, (1986) – A Multivariate Approach for Classifying Hospitals and Computing Blended Payment Rates in Medical Care, April 1986, Vol.24, No 4.
8. Zodet, M. et al, (1996) – Creation of Hospital Peer Groups in Clinic Performance and Quality Health Care,January-March, Vol. 4, No 1.